

Distributed solving through model splitting

Lars Kotthoff and Neil C.A. Moore
`{larsko,ncam}@cs.st-andrews.ac.uk`

University of St Andrews

Abstract. Constraint problems can be trivially solved in parallel by exploring different branches of the search tree concurrently. Previous approaches have focused on implementing this functionality in the solver, more or less transparently to the user. We propose a new approach, which modifies the constraint model of the problem. An existing model is split into new models with added constraints that partition the search space. Optionally, additional constraints are imposed that rule out the search already done. The advantages of our approach are that it can be implemented easily, computations can be stopped and restarted, moved to different machines and indeed solved on machines which are not able to communicate with each other at all.

1 Introduction

Constraint problems are typically solved by searching through the possible assignments of values to variables. After each such assignment, propagation can rule out possible future assignments based on past assignments and the constraints. This process builds a search tree that explores the space of possible (partial) solutions to the constraint problem.

There are two different ways to build up these search trees – n -way branching and 2-way branching. This refers to the number of new branches which are explored after each node. In n -way branching, all the n possible assignments to the next variable are branched on. In 2-way branching, there are two branches. The left branch is of the form $x = y$ where x is a variable and y is a value from its domain. The right branch is of the form $x \neq y$.

The more commonly used way is 2-way branching, implemented for example in the Minion constraint solver [5]¹. However, regardless of the way the branching is done, exploring the branches can be done concurrently. No information between the branches needs to be exchanged in order to find a solution to the problem.

We exploit this fact by, given the model of a constraint problem, generating new models which partition the remaining search space. These models can then be solved independently. We furthermore represent the state of the search by adding additional constraints such that the splitting of the model can occur at any point during search. The new models can be resumed, taking advantage of both the splitting of the search space and the search already performed.

¹ <http://minion.sf.net>

2 Background

The parallelisation of depth-first search has been the subject of much research in the past. The first papers on the subject study the distribution over various specific hardware architectures and investigate how to achieve good load balancing [13, 7]. Distributed solving of constraint problems specifically was first explored only a few years later [2].

Backtracking search in a distributed setting has also been investigated by several authors [12, 15]. A special variant for distributed scenarios, asynchronous backtracking, was proposed in [17]. Yokoo *et al* formalise the distributed constraint satisfaction problem and present algorithms for solving it [18].

Schulte presents the architecture of a system that uses networked computers [16]. The focus of his approach is to provide a high-level and reusable design for parallel search and achieve a good speedup compared to sequential solving rather than good resource utilisation. More recent papers have explored how to transparently parallelise search without having to modify existing code [10].

Most of the existing work is concerned with the problem of effectively distributing the workload such that every compute node is kept busy. The most prevalent technique used to achieve this is work stealing. The compute nodes communicate with each other and nodes which are idle request a part of the work that a busy node is doing. Blumofe and Leiserson propose and discuss a work stealing scheduler for multithreaded computations in [1]. Rolf and Kuchcinski investigate different algorithms for load balancing and work stealing in the specific context of distributed constraint solving [14].

Several frameworks for distributed constraint solving have been proposed and implemented, e.g. FRODO [11], DisChoco [3] and Disolver [6]. All of these approaches have in common that the systems to solve constraint problems are modified or augmented to support distribution of parts of the problem across and communication between multiple compute nodes. The constraint model of the problem remains unchanged however; no special constructs have to be used to take advantage of distributed solving. All parallelisation is handled in the respective solver. This does not preclude the use of an entirely different model of the problem to be solved for the distributed case in order to improve efficiency, but in general these solvers are able to solve the same model both with a single executor and distributed across several executors.

The decomposition of constraint problems into subproblems which can be solved independently has been proposed in [9], albeit in a different context. In this work, we explore the use of this technique for parallelisation. A similar approach was taken in [14], but requires parallelisation support in the solver.

3 Model splitting

We now describe our new approach to the distributed solving of constraint problems which modifies the constraint solver to modify the constraint model and does not require explicit parallelisation support in the solver.

Before splitting, the solver is stopped. As well as stopping, it is designed to output *restart nogoods* for the problem in the solver’s own input language [8]. These constraints, when added to the problem, will prevent the search space just explored from being repeated in any split model².

To split the search space for an existing model, partition the domain for the variable currently under consideration into n pieces of roughly equal size. Then create n new models and to each in turn add constraints ruling out $n-1$ partitions of that domain. Each one of these models restricts the possible assignments to the current variable to one n th of its domain.

As an example, consider the case $n = 2$. If the variable under consideration is x and its domain is $\{1, 2, 3, 4\}$, we generate 2 new models. One of them has the constraint $x \leq 2$ added and the other one $x \geq 3$. Thus, solving the first model will try the values 1 and 2 for x , whereas the second model will try 3 and 4.

The main problem when splitting constraint problems into parts that can be solved in parallel is that the size of the search space for each of the splits is impossible to predict reliably. This directly affects the effectiveness of the splitting however – if the search space is distributed unevenly, some of the workers will be idle while the others do most of the work.

We address this problem by providing the ability to split a constraint model after search has started. The approach is very similar to the one explained above. The only difference is that in addition to the constraints that partition the search space, we also add constraints that rule out the search space that has been explored already.

Assume for example that we are doing 2-way branching, the variable currently under consideration is again x with domain $\{1, 2, 3, 4\}$ and the branches that we have taken to get to the point where we are are $x \neq 1$ and $x \neq 2$. The generated new models will all have the constraints $x \neq 1$ and $x \neq 2$ to get to the point in the search tree where we split the problem. Then we add constraints to partition the search space based on the remaining values in the domain of x similar to the previous example.

Using this technique, we can create new chunks of work whenever a worker becomes idle by simply asking one of the busy workers to stop and generate split models. The search is then resumed from where it was stopped and the remaining search space is explored in parallel by the two workers. Note that there is a runtime overhead involved with stopping and resuming search because the constraints which enable resumption must be propagated and the solver needs to explore a small number of search nodes to get to the point where it was stopped before. There is also a memory overhead because the additional constraints need to be stored.

We have implemented this approach in a development version of Minion, which we are planning to release to the public after further testing and verifi-

² This same technique allows Minion to be paused and resumed: the nogoods are provided when the solver is interrupted, and can be used to restart search, potentially using a different solver, different search strategy or on a different machine.

cation. Initial experiments showed that the overhead of stopping, splitting and resuming is minimal and not significant for large problems.

In practice, we run Minion for a specified amount of time, then stop, split and resume instead of splitting at the beginning and when workers become idle. The algorithm is detailed in Figure 1. This creates an n -ary split tree of models for n new models generated at each split. Initially, the potential for distribution is small but grows exponentially as more and more search is performed.

Input : constraint problem X , allotted time T_{max} and splitting factor $n \geq 2$
Output: a solution to X or nothing if no solution has been found

run Minion with input X until termination or T_{max} ;

```

if solved( $X$ ) then
  | terminate workers;
  | return solution;
else if search space exhausted? then
  | return;
else
  |  $X' \leftarrow X$  with new constraints ruling out search already performed;
  | split  $X'$  into  $n$  parts  $X'_1, \dots, X'_n$ ;
  | for  $i \leftarrow 1$  to  $n$  do in parallel
  | | distSolve( $X'_i, T_{max}, n$ );
  | end
end

```

Fig. 1. distSolve(X, T_{max}, n): Recursive procedure to find the first solution to a constraint problem distributed across several workers.

4 Comparison to existing approaches

We see the main advantage of our approach in not requiring any involved changes to the constraint solving system to support distributed solving; in particular communications between workers. Conventionally, distribution is achieved with the aid of recomputation and cloning; established techniques used e.g. in [16]. We require two features of our solver: partitioning using constraints, and ability to output restart nogoods. Our system makes use of *cloning*, which we call “splitting” and implement by means of nogoods added to the constraint model in order to partition the domain of a variable. However, where other systems use *recomputation*, our system uses restart nogoods. In a system based on recomputation the clone begins at specific search path, e.g. stolen from another worker; with restart nogoods notionally multiple search paths are provided and the solver may explore these in any way it wishes, not necessarily one after the other. It is

merely a convenient and compact way of encoding the situation where a solver is relinquishing *all* its remaining work.

Contemporary constraint solvers make it easy to change or amend the search procedure to support distribution across several executors, but even then changes to the constraint solving system are required. While an initial implementation of distributed search can be done relatively quickly, handling failure properly and supporting things like nodes being added and removed dynamically requires significantly more effort. Our approach separates this part completely from the constraint solving system.

There are several advantages to implementing distributed solving the way given in Figure 1. First, by creating regular “snapshots” of the search done, the resilience against failure increases. Every time we stop, split and resume, the modified models are saved. As they contain constraints that rule out the search already done, we only lose the work done after that point if a worker fails. This means that the maximum amount of work we lose in case of a total failure of all workers is the allotted time T_{max} times the number of workers $|w|$.

The fact that the modified models can be stored can also be exploited to move the solving process to a different set of workers after it has been started without losing any work. It furthermore means that we require no communication between the individual workers solving the problem; they only need to be able to receive the problem to solve and send the solution or split models back.

Another advantage is that small problems which Minion can solve within the allotted time are not split and no distribution overhead is incurred. Solving proceeds as it would in a standard, non-distributed fashion.

Our approach is particularly suitable for use with existing grid-computing software or workload management systems such as Condor³. Every time new models are generated, they are submitted to the system which queues them and allocates a worker as soon as one becomes available. By leveraging existing software to perform this task, a huge amount of development time is saved and errors are avoided. For large problems, the number of queued jobs will usually exceed the number of workers, ensuring good resource utilisation.

The management system to monitor the search, queue split models and terminate the workers if a solution has been found can be implemented efficiently in just a few lines of code. We have written a Ruby script that performs this task in little more than an hour. Obviously there is potential for trying different search strategies for different branches or modifying other search parameters in order to improve efficiency. With the appropriate modifications, the management system could adapt the search procedure specifically for individual parts of the search tree. We are planning to explore these possibilities in future work.

A downside of the approach is that the number of models which can be solved in parallel will be small to start with. This means that the utilisation of resources in the beginning will be suboptimal. Only as more and more search space is explored and more and more split models are generated, the utilisation will improve. This however can be mitigated by dynamically adapting the time

³ <http://www.cs.wisc.edu/condor/>

for which Minion is run before splitting the problem – in the beginning, we set it to a small value to quickly get many models that we can solve in parallel. Then we gradually increase the allotted time as the resource utilisation improves.

Our technique is intended to be used for very large problems which take a long time (many hours, days or weeks) to solve. It is unlikely to be efficient for problems that can be solved in minutes, but on the other hand there is no need for distributed solving if the problem can be solved sequentially in a short amount of time. Only large search spaces can be split in a way that many workers are kept busy without a high communication overhead.

5 Detailed example

We will now have a detailed look at how our approach works for a specific problem. Consider the 4-queens problem. We want to place 4 queens on a 4×4 chessboard such that no queen is attacking another queen. Queens can move along rows, columns and diagonals. The constraints therefore have to forbid that two or more queens are in the same row, the same column or on the same diagonal. The constraint model in Figure 2 captures this problem.

```
language Dominion 0.1
letting n = 4
dim queens[n]: int
find queens[..]: int {1..n}
such that
alldifferent alldiff(queens[..])
diagonals1 [ not(eq1 eq(queens[i], add(queens[j], j-i))) |
             i in {0..n-2}, j in {i+1..n-1} ]
diagonals2 [ not(eq2 eq(queens[i], add(queens[j], i-j))) |
             i in {0..n-2}, j in {i+1..n-1} ]
```

Fig. 2. Model for the 4-queens problem in the Dominion language [4]. The model describes the n -queens problem in general and is specialised for 4-queens in the second line.

We assume variable ordering $queens_0, queens_1, queens_2, queens_3$, ascending value ordering from 1 to 4 and n -way branching. The search tree for a simple backtracking algorithm is depicted in Figure 3. Even for a very small problem like this, there is significant potential for distributed solving.

We now start solving the problem until we reach the assignment $queens_0 = 2$. Then we stop. The constraint we need to add to resume the search at the same point is

```
resume not(innerresume eq(queens[0], 1))
```

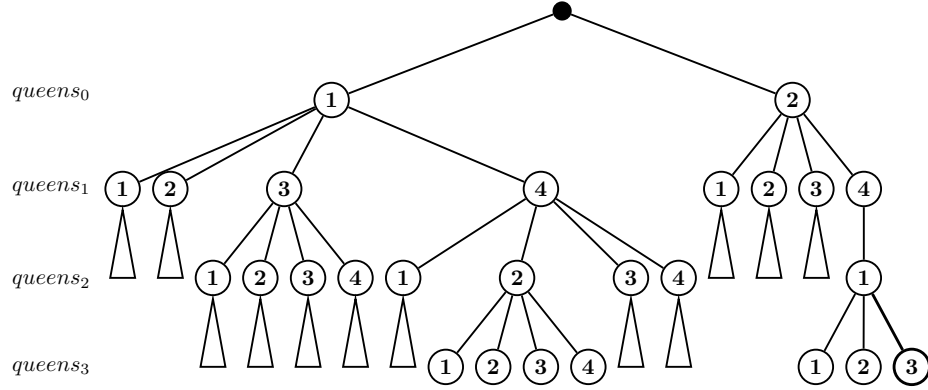


Fig. 3. First solution search tree for 4-queens. The triangles depict subtrees which are not explored because the partial assignment so far cannot be part of a solution. The bold, rightmost node is where the solution is found. The levels of the tree show assignments to the variables shown on the left.

(note that `resume` and `innerresume` are simply identifiers given to the specific constraints as required in the Dominion language [4]).

Let us assume a splitting factor of 2. We add the constraints to split the remaining search space as follows. The variable currently under consideration is `queens1`, its domain is $\{1, 2, 3, 4\}$ and therefore the constraints are

`left leq(queens[1], 2)` and `right leq(3, queens[1]).`

The search is restarted with two workers, each exploring separate branches of the remaining search space. The first worker finds no solutions in its part of the search space, terminates and returns. The second worker finds a solution and returns it. Search terminates and no further splitting is performed.

6 Conclusions and future work

We have proposed and detailed a novel approach for distributing constraint problems across multiple computers. Instead of modifying the solver to support distributed operation, we only require some simple and generic modifications that post additional constraints to the model.

The main advantages of our approach are that it does not require networked machines, is resilient against failure and can be implemented easily in constraint solvers which are aware of the state of the search.

The main drawback of this paper is that we do not have performed a systematic experimental evaluation of our approach yet. In the future, we would like to evaluate it in terms of solving speedup and resource utilisation on large, real-world problems. Furthermore, we would like to investigate finding all solutions to a constraint problem and solving constrained optimisation problems in a distributed manner.

Adapting the search procedure and parameters dynamically during search is another promising area for future work. The solving process could be tailored to the characteristics of parts of the search space to improve efficiency.

Another direction for future work is to support a higher level of abstraction for decomposing problems into subproblems. This would be necessary to support problems which cannot be decomposed by simply adding constraints that split the domain of a variable.

Acknowledgements

The authors thank Chris Jefferson for help with implementing the model splitting in Minion and the anonymous reviewers for their feedback. Lars Kotthoff is supported by a SICSA studentship.

References

1. Blumofe, R.D., Leiserson, C.E.: Scheduling multithreaded computations by work stealing. *J. ACM* 46(5), 720–748 (1999)
2. Collin, Z., Dechter, R., Katz, S.: On the feasibility of distributed constraint satisfaction. In: *IJCAI*. pp. 318–324 (1991)
3. Ezzahir, R., Bessiere, C., Belaisaoui, M., Bouyakhf, H., Mohammed, U., Agdal, V.: *DisChoco: A platform for distributed constraint programming* (2007)
4. Gent, I.P., Jefferson, C., Kotthoff, L., Miguel, I., Nightingale, P.: The DOMINION input language version 0.1. *CIRCA preprint 2009/21*, University of St Andrews (2009), <http://www-circa.mcs.st-and.ac.uk/Preprints/InLangSpec.pdf>
5. Gent, I.P., Jefferson, C., Miguel, I.: MINION: a fast, scalable, constraint solver. In: *ECAI*. pp. 98–102 (2006)
6. Hamadi, Y.: *Disolver 3.0: the Distributed Constraint Solver version 3.0* (2007)
7. Kumar, V., Rao, V.N.: Parallel depth first search. Part II. analysis. *Int. J. Parallel Program.* 16(6), 501–519 (1987)
8. Lecoutre, C., Sais, L., Tabary, S., Vidal, V.: Nogood recording from restarts. In: *IJCAI’07: Proceedings of the 20th international joint conference on Artificial intelligence*. pp. 131–136. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (2007)
9. Michel, L., Hentenryck, P.V.: A decomposition-based implementation of search strategies. *ACM Trans. Comput. Logic* 5(2), 351–383 (2004)
10. Michel, L., See, A., Hentenryck, P.V.: Parallelizing constraint programs transparently. In: *CP*. pp. 514–528 (2007)
11. Petcu, A.: *FRODO: a FFramework for Open/Distributed constraint optimization*. Technical report no. 2006/001, Swiss Federal Institute of Technology (EPFL) (2006), <http://liawww.epfl.ch/frodo/>
12. Rao, V.N., Kumar, V.: On the efficiency of parallel backtracking. *IEEE Trans. Parallel Distrib. Syst.* 4(4), 427–437 (1993)
13. Rao, V.N., Kumar, V.: Parallel depth first search. Part I. implementation. *Int. J. Parallel Program.* 16(6), 479–499 (1987)
14. Rolf, C.C., Kuchcinski, K.: Load-balancing methods for parallel and distributed constraint solving. In: *CLUSTER*. pp. 304–309 (2008)

15. Sanders, P.: Better algorithms for parallel backtracking. In: Workshop on Algorithms for Irregularly Structured Problems. pp. 333–347 (1995)
16. Schulte, C.: Parallel search made simple. In: Proceedings of TRICS. pp. 41–57 (2000)
17. Yokoo, M., Durfee, E.H., Ishida, T., Kuwabara, K.: Distributed constraint satisfaction for formalizing distributed problem solving. In: 12th IEEE International Conference on Distributed Computing Systems. pp. 614–621 (1992)
18. Yokoo, M., Durfee, E.H., Ishida, T., Kuwabara, K.: The distributed constraint satisfaction problem: Formalization and algorithms. *IEEE Trans. on Knowl. and Data Eng.* 10(5), 673–685 (1998)